FACULTY OF BEHAVIORAL SCIENCES,
UNIVERSITY OF TWENTE,
P.O. BOX 217, 7500 AE ENSCHEDE,
THE NETHERLANDS

Scientific User Interface Testing: Exploring learning effects and Fitts' Law

B.Sc.-Thesis Psychology

Cognition, Media & Ergonomics

Author:
M. Risto

1st supervisor:
Dr. E.L. van den Broek

2nd supervisors:
T. Kok, M.Sc. (CEPO beyond ICT)
Drs. F. Meijer

Table of Contents

ABSTRACT

The Scientific User Interface Testing framework (SUIT) was developed to monitor users in a natural setting and log their behavior. Color-Selector User Interface Testing (C-SUIT) is an online application of SUIT concepts in order to test color-selector user interfaces. The aim of this study was to determine whether learning takes place in the C-SUIT experiments and whether it might prevent Fitts' law from correctly modeling movement. It is assumed that pointing movements follow an 'optimal sub-movement' scheme consisting of an initial movement and an error correction phase. With practice adjustments to these phases produce a more efficient ratio between initial and corrective movements. However it is argued that Fitts' law while evolving from information theory is unable to account for learning by movement adaptation. It is hypothesized that learning effects in the C-SUIT lead to faulty predictions of movement time by Fitts' law. To this end 131 subjects participated in the C-SUIT 2.0 experiments, designed and executed by Kok (2008). In five blocks of 72 trials, five different color selectors had to be used to reproduce a color shown on a computer screen. The amount of clicks and the moved distances of the mouse were stored in a database. To search for learning effects, a correlation analysis of trial and moved distance was carried out. The data was also tested against the 'power law of practice' developed by De Jong (1957). However, no effects of motor learning were visible in the data. Further, a modeled value for the index of difficulty (ID) was used to predict the ID's of movements in the experiment. The data gave no reason to assume that Fitts' law is a good predictor of movement distance in the C-SUIT. If motor learning takes place in the C-SUIT it must reveal itself in other ways than a reduction of movement distance. The results are in stride with previous findings of Fitts' law experiments. However, the real world character of this study might have contaminated the data. Fitts' law although being able to correctly predict movement in simple pointing tasks, encounters problems when confronted with more complex tasks.

INTRODUCTION

Movement plays a central role in Human-Computer interaction (HCI). Before the Graphical User Interface (GUI) became a standard on many computers, people moved their hands and fingers over the keyboard pressing keys to give commands to the computer. In early 1980's GUI based interaction (e.g., through visual desktop environments) began to replace command line based interaction. Users learned to interact with a computer by clicking, dragging and dropping, using a computer mouse instead of keyboard commands.

With the rise of the computer from a solely working machine to a multifunctional multimedia center, it has been put to a wider use and has found its way into everyday life of millions. The ease of use became an important factor for the user. In parallel, software companies began to acknowledge the importance of usability as a feature to soar above the concurrence with ever more user friendly products.

Research in HCI may help to shift attention from a technology centered view of products to a more user centered view. For example, instead of just implementing a new powerful feature in a program, the development of usability guidelines helps people find their way more easily through that program enabling them to correctly use the feature.

Analyzing the patterns of movements that we make to achieve our goals has given researchers clues about how successful our interactions with computers are. Over time researchers figured out approaches to model these movements and predict certain factors of interaction like speed or accuracy. One of the most famous ones is the Keystroke Level Model (KLM), which gives a researcher indications on the time users take for every sub task of an interaction (Card, Moran & Newell, 1980). If a user encounters a problem while interacting with a program it is indicated by the KLM through an elevated operation time.

An important constructor of operation time is movement time. The time we need to move our mouse cursor over a certain distance to hit a target on the screen. Paul Fitts (1954) proposed a model for human movement that predicted the time it takes to move from a starting point to a target area using a pointing device. Researchers were able to expand Fitts' law so that it would also account for mouse movements in 2 dimensional computer screen environments (MacKenzie & Buxton, 1992). This should enable Fitts' Law to model mouse movements made to hit targets on a computer screen. For example, to chose a color from the Microsoft Word color selector. This is one of the main task of the so called "Color - Scientific User Interface Testing" (C-SUIT) experiments (Kok, 2008). C-SUIT was developed as an application of the SUIT framework - an online testing environment for user interfaces.

Scientific User Interface Testing (SUIT)

Over the past years research on usability led to the implementation of several usability paradigms. Kok (2008) took promising aspects of already existing interface testing paradigms and combined them with his own ideas to build a new, more powerful framework for interface testing.

Existing usability testing paradigms were subdivided in three stages. A design and an implementation phase followed by an execution phase with testing procedure and a phase for data collection and analyses. Every phase was discussed regarding its capabilities and limitations, pointing out candidates to implement in the SUIT framework. Reviewing given issues with existing designs the authors arrived at the following solutions:

Design

- To test real world interaction, it is necessary to execute the testing in real world settings instead of laboratory settings.
- Experimenter should have the possibility to implement every aspect of the interface in the testing process.
- Maximizing the number of trials keeps a certain degree of accuracy and should solve the conflict between realism for the cost of accuracy (given in lab versus real world settings).

Implementation

- With respect to location and time an approach with a home or workplace setting instead of lab setting was chosen, aiming to maximize accessibility.

Data collection

- To measure every aspect of the interaction process, Kok (2008) used both qualitative and quantitative data for evaluating the SUIT framework.

To provide a maximum amount of realism and flexibility, the interactive product is fully and in great detail simulated on a computer. This provides the experimenter the chance to test every design and measure every aspect of the interaction.

A high level of accessibility is reached through the introduction of an online testing environment which is implemented on a web server and may be accessed from anywhere through a common web browser. This approach maximizes the number of potential users. Also, no researcher is needed to interact with the participants making the experiments comparable to real life situations. Collecting quantitative data means collecting all sorts of performance data ranging from response time and mouse movement data to accuracy percentages. These measures give insight into the efficiency of the user interaction.

Qualitative data is obtained through a survey that contains questions on users' opinions on the interfaces. Through gathering various types of data, the SUIT framework facilitates a multi-view perspective. The process of combining qualitative and quantitative data, also called triangulation, is a useful method to enhance the reliability of the data by providing an additional feedback mechanism.

C-SUIT

After introducing the rules and guidelines that play a role in SUIT, the authors tested the framework in an experimental setting to see how it deals with real testing situations. The C-SUIT experiments evaluate a range of color-selector interfaces, using the new framework. Choosing among several hundred color selectors on the market, the authors picked four of them while inventing a fifth, each with a unique set of characteristics. Using the color selectors participants had to reproduce a given color that was shown on the screen. After the experiments, Kok (2008) obtained a dataset that will also be the foundation for the current analyses. In this thesis, participants' performance to the predictions modeled by using Fitts' law will be under investigation.

Fitts' law

Fitts' law (Fitts, 1954) has a successful history of predicting human movement time from a starting point towards a target using a pointing device. The model takes its basic methodology from Shannon's Theorem 17, which is used to describe the information capacity of communication channels (Shannon & Weaver, 1949). Following Shannon's formulation movements are interpreted as the transmission of units of information called bits. The number of bits assigned to a movement raises analogue to its difficulty; so, harder movements result in a higher count of bits. The human motor system processes these bits of information. The channel capacity or bandwidth of the motor system is measured by the number of bits that can be processed in a second and is called index of performance (IP). To rate and compare movements another measure, the index of difficulty (ID) is used. When modeling the execution of movements, ID is the number of bits of information that is transmitted and IP is the rate of transmission. According to Fitts IP stays constant over different values for ID. That means that the rate of human information processing does not change with the difficulty of the task. The index of difficulty can be calculated using the following formulation:

$$ID = log2\left(\frac{2A}{W}\right) \tag{1}$$

In this formula, *A* stands for the amplitude or distance from the cursors starting position to the middle of the target. *W* denotes the width of the target. That is the area people must hit to successfully terminate their task. Note that in the original formulation *W* only stand for the *width* of the target neglecting its height. Fitts (1954) originally developed his law for one dimensional tasks. This may lead to problems when predicting movement times in two dimensional tasks, on which we will elaborate later on.

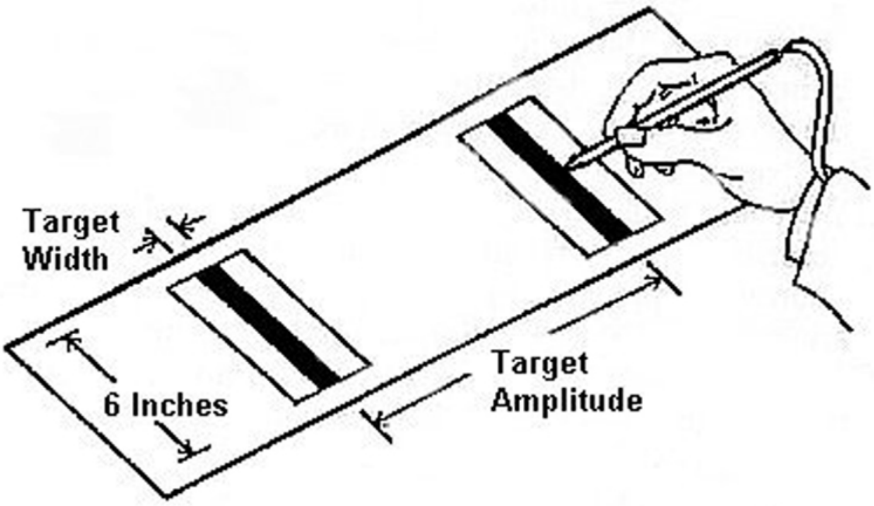

Figure 1. Fitts one dimensional tapping task

Figure 1 shows a common tapping task, as used by Fitts. The subjects were asked to tap between both target areas, using a pen as a pointing device. A common variation of this task is the discrete movement task. Here, the subjects carry out a single movement from a starting position towards a target.

The time people need for target completion (MT) can be predicted using a linear equation:

$$MT = a + b\,ID \tag{2}$$

The values for ‘a’ and ‘b’ are empirically determined constants were ‘a’ represents the initial starting time of the movements and ‘b’ stays for the speed of the movement.

For cognitive science, the translation of information into units of bits was revolutionary in two ways: 1) it introduced the bit as a new measure of task difficulty in cognitive tasks and 2) deriving from

Shannon's theorem, it offered a whole new view on information processing in humans, by introducing the channel metaphor for the human information processing system. However, this new approach bears problems that result from the fact that the human information channel is not totally equal to electronic communication channels, as described by Shannon. The description of information processing in bits per second is straight forward according to electric information channels. However, it may lead to faulty conclusions for the use in human information processing experiments. Sanders (1998, p. 12) argues that information theory only describes input-output relations but neglects internal processing mechanism and feedback. That might work for static tools as telecommunication systems but not for humans because they are capable of learning and modulating their input-output relations with practice.

Fitts' law in HCI

The application of Fitts' law spreads from kinematics to human factors. From the late 70s on, it has had impact on the field of human computer interaction (HCI). It was mentioned earlier that a task, analogue to the discrete taping task from the original experiments by Fitts, can be transposed into interactive computer tasks, using a monitor and a cursor operated by a mouse or a mouse-like pointing device. Researchers use his formula to predict the time a person needs to move a mouse cursor from a starting position to a target location somewhere on the screen given the width of and distance (this distance is sometimes called amplitude) to the target.

The first scientific application of Fitts' law in HCI was by Card, Englisch and Burr (1978). They evaluated four different input devices (mouse, joystick, step keys, text keys) comparing how efficient they can be used to select text on a CRT display. The results were also tested against predictions from Fitts' law with the help of regression analysis. The mouse was found to be the fastest, while producing the lowest error rates. In this study, Fitts' law accounted for both the mouse and the joystick movement times.

Whisenand and Emurian (1999) used Fitts' law to analyze the variance of movement times in both a pointing task and a drag and drop task using targets of differing size and shape. Reaction times obtained in the experiments were compared to the predictions from Fitts' law. They concluded that for an optimal speed and accuracy tradeoff displayed targets should be accord with the following guidelines. Target objects should be square shaped sized between 8 and 16 mm. They should be located 40 mm or less from the starting point and they should be approached from a horizontal or vertical angle. This is a good example of how Fitts' law could help Software developers with the development of interfaces by telling them how to structure and size menus or icons on the screen in a way that people can interact with them faster and more accurate.

Criticism of Fitts' law

Although the model is based only on target width and amplitude, it "has proven one of the most robust, highly cited, and widely adopted models to emerge from experimental psychology" (MacKenzie, 1992). Despite its appeal, Fitts' law has been the target of criticism from researchers. MacKenzie (1992) summarized the issues researchers had to deal with, while applying Fitts' law to HCI questions. He also argues that the large amount of inconsistencies across Fitts' law studies makes them difficult to compare.

One of the problems is that Fitts' law failed to accurately predicted reaction times for low values of ID.

In several studies, (e.g., Welford, 1960; Meyer et al., 1982) scatter plots of reaction times revealed an upward curvature away from the regression line when ID approached zero.

Numerous researchers have made suggestion for improving Fitts' law. According to MacKenzie (1992), the Welford variation (Welford, 1960; 1968, p. 147) is the most widely adopted:

$$MT = a + b\ log2\left(\frac{A}{W} + 0{,}5\right) \qquad (3)$$

MacKenzie (1992) also proposes his own variation of the formulation. He points out that both the original and the Welford formulation sometimes yield unrealistically low and even negative values for ID. This is a theoretical problem for researchers that are left with the difficult task of explaining the meaning of negative ID values. His new formulation always gives positive values for task difficulty.

$$MT = a + b\ log2\left(\frac{A}{W} + 1\right) \qquad (4)$$

It has been said earlier that Fitts’ law initially was developed for one dimensional tasks. In other words, subjects approached the target from only one direction with target amplitude and target width measured along the same axis. Only the width of a target area had to be included in the formulation. This leads to issues when applying Fitts' law to predict movement time in two dimensional pointing tasks. As the target can be approached from different angles the ratio between width and height of the target gains importance. In particular problems emerge with rectangular targets as W changes with approach angle. A small but wide target would yield a higher value for ID when approached vertically and lower for ID’s when approached horizontally. Finally, several models for the computation of target width emerged.

The use of the horizontal extent of the target as target width is called the status-quo model because it was most widely adopted (MacKenzie, 1992). However, this model, because of the aforementioned reasons, sometimes yields negative estimates for task difficulty. The alternative W' (spoken: W prime) model measures target width along the targets approach axis. Yet another model, the smaller-of model, takes the smaller value of target width or height for W. In an experiment, the smaller-of model and the W' led to significant better predictions of reaction time than the status-quo model (MacKenzie, 1995).

The current research

Research has shown that aiming movements consist of two phases. An initial impulse followed by an error correction phase (Meyer et al., 1988; Abrams, Meyer & Kornblum, 1990). The initial impulse is characterized by a rapid limb movement in target direction that stops when the target is reached. If the endpoint of this first impulse does not hit the target, movement will be adjusted in the error correction phase. This phase is said to consist of on-line, sensory based adjustments to movement in order to reduce the discrepancy between pointer and target. It should be noted that the initial impulse phase, does not contain movement modifications and is comprised of one sub movement. The error correction phases may consist of either a single or multiple sub-movements (Crossman & Goodeve, 1983; Meyer, Smith, Kornblum, Abrams, & Wright, 1990).

Meyer et al. (1988) found that total movement times can be minimized, while maintaining a high degree of accuracy, by achieving an optimum in the trade-off between the durations of the initial impulse and error correction phase. On the one hand, this means that initial impulses at a high velocity have short durations but frequently miss the target region. So, afterwards the increased need for error correction will increase total movement time, even though the initial impulses have short durations. On the other hand, initial impulses at low velocities are highly accurate; but, the long initial impulse durations will result in long total movement times.

Gottlieb et al. (1988) did conduct research on fast elbow flexions. They showed that over time, subjects learned to execute movements faster over the practiced distance by refining their neural commands to the muscles. So, possibly even movement in simple tasks can be affected by learning. This led to the assumption that by adjusting their neural commands, participants would achieve maximum performance. This would result in an as high as possible initial impulse velocity, combined with a minimized error correction phases. By minimizing corrective movements, the overall distance, a crucial factor for Fitts' model, also reduces. So the question is:

Do reaction times show learning effects over single trials in the C-SUIT experiments?

Card, English and Burr (1978) showed that learning occurred in between trials of their pointing task. They were able to demonstrate, that a learning curve of positioning time versus amount of practice can be modeled using the formula developed by De Jong in 1957. Earlier it was argued that because Fitts' law evolved from information theory, it neglects learning effects while describing only untrained movements but not movements that are executed after some time of practice (Sanders, 1998, p. 12). Based on the aforementioned theories about motor learning, the findings by Card, English and Burr and the shortcomings of information theory, it is assumed that:

Fitts' law, while neglecting learning effects, is unable to model movement in the C-SUIT experiments.

It should be noted that the SUIT framework provides the possibility for data triangulation. Both qualitative and quantitative data were obtained in the C-SUIT. Having the this data collection, it might be interesting to see whether or not possible learning effects take effect from third factors as the evaluation of the color selector interface or the amount of computer experience.

METHOD

The data used in the current research is derived from the database obtained by Kok (2008) in the C-SUIT 2.0 experiment. For this reason, a brief description of the design of this experiment will be provided.

Participants

A total of 131 individuals, of which 38 men and 93 women, took part in the experiment. They were aged from 18 to 40 years, with a mean age of 21.18 years (SD: 2.62). 122 were right handed and 9 were left handed. All participants hand normal or corrected to normal vision, with 55 of them wearing glasses or contact lenses. The majority (121) of the participants came from a University setting, 9 were high-school students and 1 had higher education. 48 reported to have no experience with colors, 78 had some and 9 reported much prior experience. All participants reported to have two or more years of computer experience; the majority (76) had between five and ten years of experience. Upon registration it was determined whether or not participants were color blind; if this was the case, they were denied participation. Table 1 provides an overview of the different categories the participants can be grouped in.

Table 1. The participants for the C-SUIT 2.0 experiment, adopted from Kok (2008).

Gender	Male: 38, Female: 93
Education	High school: 9, Student: 121, Higher education: 1
Handedness	Left: 9, Right: 122
Glasses/lenses	Yes: 55, No: 76
Color experience	None: 48, Some: 78, Much: 9
Age	Average: 21.18, SD: 2.62, Min: 18, Max:40
Computer experience	2 yrs: 5, 3yrs: 7, 4yrs: 13, 5-10yrs: 76, >10yrs: 34

Apparatus

Subjects were able to conduct the experiment with any computer that fulfilled the basic system requirements. This makes it impossible to give a more detailed description of the used equipment. According to Kok, both laptops and desktop computers were used as well as both CRT monitors and TFT flat panels. To be used in the experiments a computer had to meet the following requirements:

- A pointing device; e.g., a computer mouse, a trackball or a touch pad. However, a mouse was recommended.
- Internet connection in order to get in contact with the server. Because of the large amount of data transmitted, especially from the mouse movements, a broadband connection was preferred.
- A web-browser was needed that supported Java-Script.

All requirements were checked automatically before the experiment was started. Participants that did not initially meet the requirements were given advice to solve any incompatibilities.

Screen information (i.e., resolution and color depth) as well as browser data (i.e., manufacturer and version) were recorded, while they could influence the experiment outcome. The software used was part of the SUIT framework developed by Kok (2008) and colleagues.

Participants were required to complete the task, using the following five color selector interfaces given in table 2:

Table 2. **Color selector interfaces used in the C-SUIT**

Interface	Name
1	Color 30
2	Color picker
3	MS Paint
4	MS Power Point
5	VindX

These user interfaces were selected in order to obtain a high level of diversity between them. Their layout varies when it comes to color availability and usability.

Design

The experiment has a within subjects repeated measures design. The independent variable is the type of color selector used. In order to achieve triangulation, additional independent variables are the prior experience with colors and computers and the rating of the color selector by the users, as obtained through a post experimental survey. The dependent variable in the experiment is movement time (MT), defined as the time between the appearance of the user interface on the screen and the final color selection marked by a button press in a corner of the screen.

With the current analyses, the index of difficulty (ID) as dependent variable will be used, instead of MT. This does not make a difference, as we take a look at Formula 4. The values for *a* and *b* are constants; so, the only changing factor is ID. Formula 1 shows that ID is computed from the width and the amplitude of the target button. For every user interface, target width remains the same. Target amplitude defined as the distance from the starting position to the target, therefore, is the only changing factor.

Measuring learning as a change in target distance implies that first an optimal distance from the starting point to the target has to be known. The distance can easily be measured since it is the straight line from the start (point A) and the target (point B).). It is then assumed that the actual movement of the mouse from point A to B does not equal a straight line. Whisenand and Emurian (1999) observed that participants movements in a pointing task varied with target distance and width. Especially with long amplitudes their cursor movements were sometimes described as 'looping' motions towards the target. Others were executing long rapid movements in the general direction followed by smaller, more precise movements. This pattern is similar to the two phases hypothesized by Meyer et al. (1988). It is assumed that over time the curved line of movement will straighten out as participants make fewer errors, while moving from point A to point B. Figure 2 shows the initial and the assumed final phase of this learning process. Over time, the dotted line starts resembling the straight line as participants make fewer errors choosing a color.

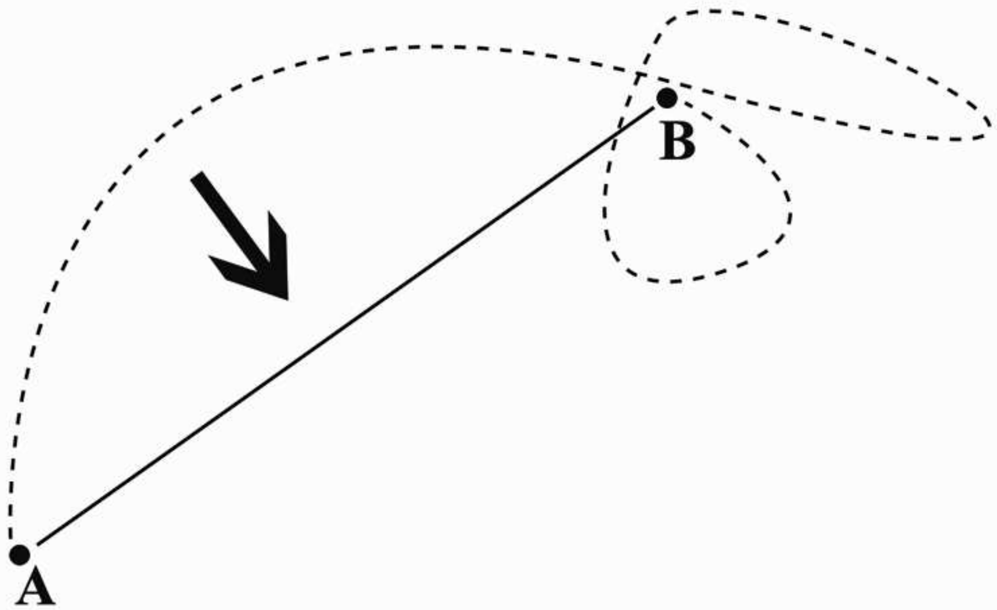

Figure 2. Learning defined as a reduction of moving distance. With time and practice, mouse movements (dotted line) are expected to resemble a straight line from A to B.

Task & Procedure

Participants have to visit a website to take part in the experiment. In line with the premises of the SUIT framework, the whole experiment takes place through this website. In the beginning of the experiment, a welcome screen is shown that provides the user with an overview of the procedure. Both background information and instructions are provided to the participants. The participants are also asked to register. In this registration process, they give information about themselves and their color and computer experience. Upon registration every participant is assigned a randomized sequence of the five blocks of trials. Every block represents a GUI and every trial represents one color selection loop. The order of the color selector interfaces is counterbalanced and each participant is randomly assigned a sequence of interface and color stimuli. Those color stimuli are a random sample set of the computers full palette of 16.777.216 (256^3) colors.

At the beginning of a loop, a rectangle is placed on the screen filled with the stimulus color. The colors are briefly (less than a second) shown. Each color is represented by a filled rectangular shape, surrounded by default (gray) background color. After the stimulus color disappeared, the color selector is revealed awaiting action for the user. Participants use a pointing device to pick a color from a presented color selector in order to reproduce a color previously seen on the screen.

After a color is picked by the user, the system records the time and accuracy of the color selection. In order to keep the subject apprised of its progress, a small indicator was added to the lower left, showing how many stimuli remain. A small break between at the change of interface, allows the subject to rest and provides an option to suspend the experiment. Because participants are allowed to participate in an environment of choice, they are given the possibility to pause the experiment at any given time. Later, they can re-enter the session at the point where they had paused it. This gives participants the possibility to fully attend to the experiment and take a break when their attention is being deviated.

When the color selection procedure is completed, the survey is loaded. A list of questions deals with the subject’s experiences of each selector, asking for the interface he or she liked most (with arguments) and which interface he or she thinks was most accurate.

RESULTS

Data preparation

In order to do proper analysis on the data, it had to be modified to meet the experiments requirements. From the SUIT2 database, it was possible to obtain whole movement distances per single trial. As mentioned previously, in the SUIT2 experiment there was no single universal starting position from which the user would start moving the mouse. In fact the user could move his mouse anywhere between trials after he had pressed the “Further” button resulting in a unique and unpredictable starting point every trial. This is different to the methods used in previous experiments that tested Fitts’ law (e.g., MacKenzie & Buxton, 1992; Whisenand & Emurian, 1999). In these experiments, the user had fixed starting and ending points for every block of trials; so, the distance between both points could be controlled. Although the SUIT2 data shows that few users made excessive mouse movements before trials, it is a point that should be considered when analyzing the data.

A second difference to previous Fitts’ law experiments was the number of clicks that were possible in one trial. After making their first click on a chosen color, users were allowed to change their minds and chose another color if they thought that this would better fit the reference color. In some cases this resulted in up to 48 mouse clicks in one single trial. While imagining that between every click the user may had a short time to compare colors and think about their choice this would result in unrealistic high values for movement time per trial. In order to account for multiple mouse clicks and changing starting positions, a reference distance had to be computed. It was assumed that every time a user clicked the mouse a color was selected. Adding up the minimum distances between separate mouse clicks yielded the minimum overall movement distance. Further, by subtracting this minimum movement distance from the measured overall distance, the relative movement distance for every trial could be received. The relative distance was filtered for values between 1 and 10 clicks. Finally, distance values under and above 3 SD were excluded from the analyses. Then, the data was split over the five user interfaces. Finally, mesured relative distances ranged from 4 to 1834 pixel and had a mean of 777 pixel (SD 404 px.).

Testing for learning effects

In the previous chapter it was assumed that learning would be visible in a reduction of movement distances as the number of trials increased. As an indication of a general effect of learning, a moderate negative correlation between trial and movement distance should be visible. However, a Pearson correlation of trial and the relative distance shows now learning effects. A moderate correlation would be a between r = -0.3 and r = -0.5. However in our analysis we were not able to achieve even a small correlation around r = -0.1. Table 3 gives a comparison of R for the five GUI's.

Table 3. **Correlations between trail and distance**

Interface		Relative distance
1	**Trial**	-0.026*
2	**Trial**	-0.044**
3	**Trial**	-0.041**
4	**Trial**	-0.014
5	**Trial**	-0.061**

Note. *p < .01, **p < .001

A scatter plot was made to see if the data would fit another pattern, different from a negative correlation. Figure three gives an example of the distribution of distance values per trial for user interface one (Color 30). What can be seen here is the random distribution of distance values as the number of trials increases. The visible horizontal grouping of distance values that can be seen is an effect of the different screen resolutions used in the SUIT2 experiment. Higher resolutions resulted in longer ways as more pixel had to be covered with the mouse movements.

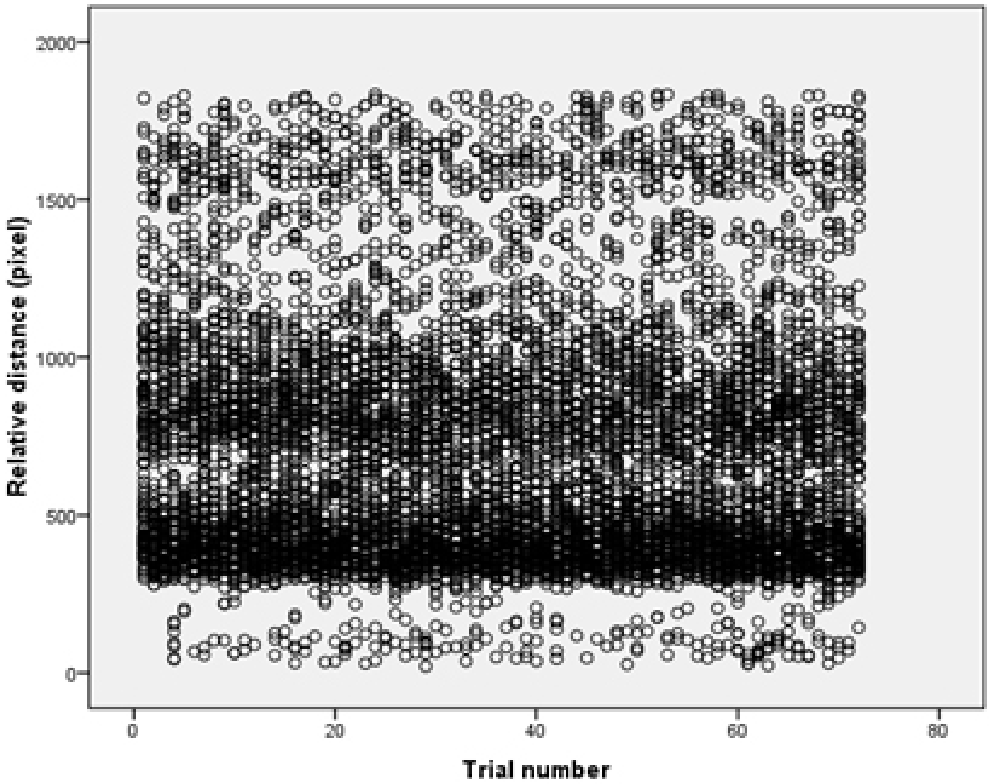

Figure 3. Scatter plot of relative distance and trial number, adopted from Kok (2008).

Although the data showed no direct visible learning effects an attempt was made to compare it to a learning model. Card, English and Burr showed in 1978 that their data, obtained by using several pointing devices, would show improvements with practice and fit a learning curve approximated, using the formula by De Jong (1957). Applying the model to the data a learning curve can be obtained. The values from the curve can then be used to make a prediction about the data.

In the late fifties De Jong found that the logarithm of the time taken to achieve a particular task decreased linearly with the logarithm of the number of trials. This observation later became known as "the power law of practice". De Jongs formulation gives movement time as a function of the amount of practice.

$$T_N = T_1 N^{-a} \tag{5}$$

Looking at MacKenzie's (1992) variation of the Welford formulation (see formula 4), we see that movement time is calculated, using distance (A) to and the width (W) of the target. The values for 'a' and 'b' are constants that are determined empirically. Because per GUI also the target width stays constant the only varying values is target distance. So it should make no difference to apply De Jong's formula using only the moved distance.

$$D_N = D_1 N^{-a} \tag{6}$$

Here

D_1 = Stands for the traveled distance on the first trial

D_N = Stands for the traveled distance on the Nth trial

N = is the trial number

α = is an empirically determined constant

The ease of learning with each GUI can be described by the two numbers D_1 and α. We can determine both numbers by fitting formula 6 to a straight line. While taking the log (base 10) on both sides of formula we will receive:

$$\log D_N = \log D_1 - \alpha(logN) \tag{7}$$

Now, we can determine D_1 and α by regressing $logD_N$ on logN.

Through making use of the power law of practice we obtained a modeled learning curve and predicted values for movement distance (D_N) for every trial. Table 4 gives the values predicted by equation 7 along with the squared multiple correlations from the regression of the relative distance on D_N. Alpha (α) is an indicator of how practice improved movement distance in all five interfaces. As we can see per GUI, there are differences in the improvement with practice although the overall effects of practice are very small. We can also see from the regression analysis that the learning model by De Jong was not able to predict the changes in relative movement distance over trials. Equation 7 explains only 4% of the variance in the average movement distance for interface 3 (MS Paint color selector) and even less of the variance in the other interfaces.

Table 4. **Learning Curve Parameters**

Interface	**D_1 (pixel)**	**α**	**Learning Curve Equation**	**R^2 from regression with relative distance**
1	732,97	0,028	$D_N = 732{,}97 * N^{-0{,}028}$	0,028
2	870,79	0,021	$D_N = 870{,}79 * N^{-0{,}021}$	0,021
3	779,92	0,036	$D_N = 779{,}92 * N^{-0{,}036}$	0,041
4	797,81	0,01	$D_N = 797{,}81 * N^{-0{,}01}$	0,013
5	841,29	0,054	$D_N = 841{,}29 * N^{-0{,}054}$	0,011

Testing of Fitts' law against the data

To test the predictions by Fitts' law, modeled values for ID were used to predict the real ID's of movements in the color selectors. Instead of using an overall value for ID per color selector, for every trial the modeled distance was used to compute the modeled ID for that specific movement using formula 1. It should be noted that this approach is different from earlier approaches in the literature. In an earlier study, Van den Broek, Kisters and Vuurpijl (2004) used Euclidean distances to compute the internal ID of their color selectors. This resulted in a standard distance for every color selector. Although this approach is very elegant, it assumes that the starting position for a pointing task is in the center of the color selector. This was not true for the C-SUIT experiments, where the starting position of the mouse could be inside or outside the color selector. In this new approach, instead of computing a single standard amplitude for every color selector, for every trial the minimum distance between the mouse clicks was used instead. The sum of the shortest distance values between clicks equals the minimum distance per trial. An example of a trial with four clicks is shown in figure 4. This procedure would result in a more accurate prediction of the ID of movement in a given trial while allowing more than one mouse clicks even outside the interface.

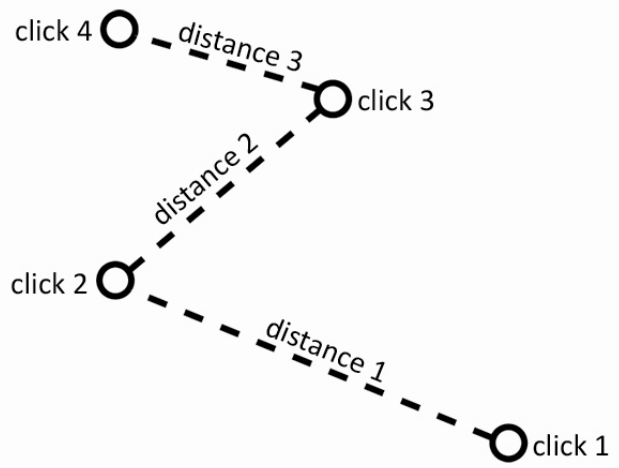

Figure 4. The minimum distance equals $\sum$distance.

To obtain the experimental movement ID's the relative values for distance were used to compute a relative ID. A regression of relative ID on modeled ID is shown in Table 5.

Table 5. **Regression of relative ID on model ID**

Interface	R	R^2	Standard error
1	0,24[a]	0,057	0,78
2	0,19[a]	0,036	0,72
3	0,259[a]	0,067	0,77
4	0,198[a]	0,039	0,78
5	0,275[a]	0,076	0,71

a. Predictor: Model ID

R gives the correlation of modeled ID and relative ID while R^2 gives the percentage of variance in the relative ID that can be described by the modeled ID. As we can see the predictive value of model ID is rather low explaining only up to 7,6 percent of the variance in interface 5 (VindX). From the data we have no reason to assume that Fitts' law is a good predictor of the movement distance in the C-SUIT experiments.

DISCUSSION

The purpose of this study was to search for effects of motor learning in the C-SUIT2 experiment and to further test Fitts' law ability to model mouse movements over trials. In the introduction, the hypothesis was defined that motor learning would take place as the number of trials increased. This was tested by two different approaches.

Learning was determined as the reduction of the relative distance that the mouse moved. First, a correlation analysis yielded a very small negative correlation between movement distance and trial number. This finding lead to the conclusion that learning between trials did not have a significant influence.

Next, a learning curve from the data was approximated, using the formula by De Jong (1959). It was used to predict the reduction in movement distance over time. Regressing the relative distance on the modeled distance, the learning curve had no predictive value for the C-SUIT2 data. Hence, it seems safe to assume that no motor learning has taken place over trials in the C-SUIT experiments.

The results are in contrast with Meyer et al. (1988), who proposed that participants would adjust their neural commands to the limbs and by that optimizing the relation between their initial and corrective movement phases. A possible explanation for this contradiction could be the amount of movement that has taken place between the mouse clicks. Participants could have hovered over colors before clicking them. Also the way they went from their final color to the "further" button may have contaminated the distance values.

Gottlieb and colleagues (1988) stated that practice improves even simple movements. However, also in their research the same simple movement patterns were repeated over and over again. This repetition would result in a learning process, as an adjustment in both phases of movement. In contrast, the movement patterns in C-SUIT were not repetitive, because of their ever changing directions and the sometimes repetitive clicks. This means that every adjustment made in one trial will meet a different angle and distances in the following trial. Because of the ever new movement, no single optimized pattern can facilitate all movements. It is possible that the changing movement patterns kept participants in the C-SUIT experiment from adjusting to one movement pattern even as the amount of trials increased.

Alternatively it can be argued if learning would show itself in the reduction of movement distance. An explanation for the absence of any learning effects would be that the learning solely resulted in an acceleration of the mouse movement and not in a decrease of movement distance. Since distance was used as the dependent variable, and not movement time, this research did not account for learning that resulted in movement acceleration. The 'optimized-submovement' model may not be able to explain movement acceleration in the C-SUIT as the optimization would be of a different kind. Participants would start slow towards the targets, making sure they hit the target. As they get comfortable with the task in later trials, they accelerate their movements while still maintaining a low error rate and by that keeping their second movement phase short.

Another possibility is that learning only takes place in the earliest trials. The previous analysis looked at a relative large amount of 72 trials per participant. However another correlation analysis of only the first ten trials and relative distance yielded comparable values for R. Only on interface five (the VindX color selector) a small negative correlation ($r = -0{,}13$; $p < 0{,}005$) was visible.

Whisenand and Emurian (1999) argued that Fitts' law was most accurate in modeling movements aimed at targets larger than 8mm) over a distance shorter than 40mm. Accuracy declined as the distance between starting position and target increased. These demands are rarely met in real world user interfaces. The size of the color selectors used in the C-SUIT ranged from 1 to 25 pixel making predictions from Fitts' law questionable.

Additional analyses were planned. Possible effects due to the number of colors and computer experience as well as the affective evaluation of the GUI could have interacted with learning effects. However, since no learning effects were found, these analyses were not carried out.

The second hypothesis was that, because of learning effects, Fitts' law would fail to give a correct prediction of ID for the color selectors used in the C-SUIT experiment. The modeled ID (MID) values were tested against the relative values for ID (RID) from the experiment. First, a regression analysis showed that Fitts' law was not able to predict the IDs of the interfaces tested. Also, a residual analysis revealed no equalities in mean MID and RID. This result, although confirming the second hypothesis, is misleading. The lack of a learning effect means it cannot account as an explanation for the lack of correct predictions made. The results must have been caused by other reasons.

The current analyses used a different approach than previous research did. Earlier experimenters who were able to proof Fitts' law used easier approaches, without multiple mouse clicks, avoiding multiple adding values for amplitude. It seems that the same circumstances that refrained from measuring movement distances, also refrained Fitts' law from working. This leads to the conclusion that, although Fitts' law was unable to predict the real movement IDs, it was not because of learning effects

but may have been the cause of several other reasons adding up to influence the measure.
It should also be noted that because this research was carried out after the C-SUIT experiment had been carried out; no attempt for adjusting the methods was possible. So, it should be stated that conditions could not be perfectly attuned to this experiments needs. Unfortunately, it was not possible to divide the participants total moved distance per trial on the mouse clicks made in this trial. This would have made an analysis of the initial distance before the first click possible and had pushed the experiment more in the line of traditional Fitts' law experiments. This result might lead to the assumption that Fitts' law although being able to correctly predict movement in simple pointing tasks, encounters problems when confronted with more complex tasks.

Research has still to prove whether or not Fitts' law holds for pointing tasks in real world settings; e.g., adding mouse clicks. Moreover, further research is needed on possible learning effects on movement times. It would be interesting to look into the relation of speed increase and/or distance decrease over trials. Also, future research has to determine what effect changing starting and target positions have on learning motor patterns and on the adjustment made on initial and corrective movement phases.

REFERENCES

Abrams, R., Meyer, D.E., & Kornblum, S. (1990). Eye-hand coordination: Oculomotor control in rapid aimed limb movements. *Journal Experimental Psychology: Human Perception and Performance, 15*, 248-267.

Broek, E. L. van den, Kisters, P. M. F., & Vuurpijl, L. G. (2004). Design Guidelines for a Content-Based Image Retrieval Color-Selection Interface. In B. Eggen, G. van der Veer, and R. Willems (Eds.), *Proceedings of the Conference on Dutch directions in HCI, ACM International Conference Proceeding Series.* June 10, The Netherlands - Amsterdam. New York, NY, USA: ACM Press.

Card, S.K., English W.K. & Burr, B.J., (1978). Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics, 21(8)*, 601-613.

Card, S.K., Moran, T.P., & Newell, A. (1980). The Keystroke-Level Model for User Performance Time with Interactive Systems. *Communications of the ACM, 23*(7), 396-410.

Card, S. K., Moran, T. P., & Newell, A. (1986). The model human processor: an engineering model for human performance. *Handbook of Perception and Human Performance, 2,* 45-1-35. New York: Wiley.

Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. *CHI '01 extended abstracts on Human factors in computing systems*, 281-282, New York, NY, USA: ACM Press.

De Jong, R.J. (1957). The effects of increasing skill on cycle-time and its consequences for time-standards. *Ergonomics, 1(1)*, 51–60.

Fitts, P.M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology, 47*, 381 – 391.

Gottlieb, G.L., Corcos, D.M., Jaric, S. & Agarwal, C.G. (1988). Practice improves even the simplest movements. *Experimental Brain Research, 73*, 436-440.

Grossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Pitts' law. *Quarterly Journal of Experimental Psychology, 35A*, 251-278. (Original work presented at the meeting of the Experimental Psychology Society, Oxford, England, July 1963).

Kok, T. (2008). SUIT - Scientific User Interface Testing. Unpublished M.Sc.-thesis. Donders Center for Cognition & Institute for Computing and Information Sciences, Radboud University Nijmegen.

MacKenzie, I. S., & Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '92*, 219-226. New York: ACM.

MacKenzie, I. S. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction, 7*, 91-139.

MacKenzie, I. S. (1995). Movement time prediction in human-computer interfaces. In R. M. Baecker, W. A. S. Buxton, J. Grudin, & S. Greenberg (Eds.), *Readings in human-computer interaction (2nd ed.),* 483-493. Los Altos, CA: Kaufmann. [reprint of MacKenzie, 1992].

Meyer, D.E., Smith, J.E.K., & Wright, C.E. (1982). Models for the speed and accuracy of aimed limb movements. *Psychological Review, 89*, 449-482.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review, 95*, 340-370.

Meyer, D. E., Smith, J. E. K., Kornblum, S., Abrams, R. A., & Wright, C. E. (1990). Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. *M. Jeannerod (Ed.), Attention and performance XIII* (pp. 173-226). Hillsdale, NJ: Erlbaum.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill.* Mahwah, NJ: Lawrence Erlbaum.

Soukoreff, R. W. & MacKenzie, I.S., (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal Human-Computer Studies, 61,* 751–789.

Weaver, W., & Shannon, C.E. (1949). *The Mathematical Theory of Communication,* Urbana, Illinois: University of Illinois Press.

Welford, A.T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics, 3*, 189-230.

Welford, A.T. (1968). *Fundamentals of skill.* London: Methuen.

Whisenand, T.G., & Emurian, H.H. (1999). Analysis of cursor movements with a mouse, *Computers in Human Behaviour, 15*, 85-103.